Attachment 1

Draft Statement of Work

Background

This contract supports a continuing project of data integration in the Office of Compliance (OC) which is a part of the Office of Enforcement and Compliance Assurance (OECA). The OC data integration project, called IDEA (Integrated Data for Enforcement Analysis), provides software that integrates data from many programmatic databases. IDEA is used to catalyze better enforcement (through improved targeting, case screening, and case preparation), support pollution prevention, provide for analyses of conditions responding to environmental justice concerns and studies, and participate in ecosystem protection. IDEA also provides a variety of tools such as aggregate data, derived data, and special reports that contain a variety of analyses in support of enforcement and Agency initiatives. Data integration, better enforcement, pollution prevention, environmental justice, risk assessment, ecosystem protection are all priority EPA goals.

Features of IDEA and the data integration effort are to respond to changing requirements that involve new databases, new analytical methods, and development of interfaces and query tools that provide functionality and use of the system to a wide range of users, both inside and outside the Agency. The IDEA project attempts to normalize the data where feasible to provide better understanding and use of the data across the different media. Within the scope of the contract, elements of assessment of risk from release and exposure to toxic chemicals and other analytical tools will be explored as part of OECA's effort to focus resources on the most harmful threats to human health and the environment.

OC is continuing to integrate more data, especially data beyond compliance status and compliance history: for example, release and emissions, demography, toxicity, exposure, risk, sensitive populations, ambient and biology data. The enhanced set of data supports EPA and state decisions on enforcement targeting and pollution prevention in particular, and allocation of scarce resources in general. Another effort is to make the application and data readily accessible by the public through the development of Web interfaces which allow the user to build queries and review reports over the Internet. The current Web interfaces are the Online Tracking Information System (OTIS), an Intranet based tool available to EPA and States for enforcement planning and targeting, and Enforcement and Compliance History Online (ECHO), a public view into IDEA data. Both OTIS and ECHO generate queries for the IDEA data base.

OC uses models, graphics, mathematical transformations, and expert interpretations of the integrated data to create derived and aggregated data for management reporting. OECA intends to

continue its efforts to improve the query interfaces and the user's ability to correctly interpret and make use of data displayed on reports. OC also continually improves the performance of the mainframe retrieval software and the IDEA interfaces to better answer user needs.

Finally, the complexity and imaginative possibilities of programs that integrate vast amounts of data in significant analytical depth require that a support staff be available to help users. OC wants to offer assistance, not only in the mechanics of using the integration software, but also in explaining the meaning of the data to make it easier for IDEA users to formulate questions.

Scope of Work

The contractor shall provide support in five areas. The areas are: (1) Original Software Design and Development and Existing Software Enhancement, (2) Software Maintenance, (3) Database Maintenance, Addition of New Databases and Database Surveillance, (4) User Interfaces and Tools and (5) Identification of Data and Methods for Analysis, including Risk Assessment and Modeling.

(1) Original Software Design and Development and Enhancement of Existing Software

The contractor shall devise and write computer programs that are components of the IDEA engine used to process queries and retrieve data for use by EPA staff and others in enforcement targeting, case development and multimedia, cross-program enforcement. This software will involve original work in interactive retrieval and analyses of data from many national databases, currently 30 in number, but likely to increase. The effort is to be undertaken with attention to the efficient management of main and auxiliary storage and to efficient performance, features required in an interactive system, and an especially important consideration in an environment of many large databases. The process of creating the computer programs will involve designing experiments, the results of which will support OECA's evaluations of the difference in performance among competing approaches.

Work will continue on the extension and on the addition of mainframe user interfaces, both graphical and non-graphical, and on improved retrieval strategies and algorithms.

The principal work is to be done on a mainframe computer, (currently EPA's IBM 9672 located in Research Triangle Park, NC, and used from the Washington area via telecommunications lines). Currently, the principal programming language is PL/I; assembly language is used where PL/I does not perform at all or performs inefficiently. Telecommuni-cations transport layer protocols are APPC and TCP/IP; content layer protocols are IDEA and HTTP/HTML. Client programs use ISPF, Perl, PL/I, assembly and some C. Other software could be added as determined by EPA to be necessary. The contractor shall be expected to exploit the power of parallel processing, the vector processor, when appropriate, dataspaces and hyperspaces, the use of address spaces above 16 meg, the novel use of the computer instruction set, and any other feature of the computer which will benefit complex data retrieval and analysis in an interactive setting, and as determined to be appropriate by EPA.

(2) Software Maintenance

There is a body of existing software designed and implemented by OECA according to the design features supporting interactive software described in paragraph (1). The contractor shall be familiar with this software and be prepared to make timely changes in it as bugs are uncovered, if the source data changes in substance or in format, or any other factor arises which compromises or limits the ability of IDEA to perform satisfactorily.

(3) Database Maintenance, Addition of New Databases and Database Surveillance

The IDEA system contains copies of the original programmatic databases which are refreshed regularly, usually monthly. The contractor shall provide technical support to EPA personnel in performing the monthly refresh and verifying and testing that the refresh process was successful by applying appropriate statistical measures and examination of the resulting data. Refresh software currently principally uses PL/I, assembly language, COBOL, REXX, SQL, Natural. Other languages are added as needed to accommodate source databases. Refresh automation is established using JOBTRAC.

Based on enforcement and Agency priorities, additional databases shall be added to IDEA to provide data for more comprehensive analyses. These databases may involve enforcement and compliance information, ecosystem and environmental data, demographic and disease data, or worker exposure data. The contractor shall diagnose the data structure, propose and justify the data skeleton in IDEA, write programs that will extract and copy data into IDEA, and prepare the necessary procedures for handling regular refreshes.

Original databases periodically undergo enhancements or redesign necessitating a constant surveillance to assure that IDEA faithfully replicates the data. Regular contacts with system managers are required to discuss changes and enhancements planned for the program databases and to plan for corresponding changes in IDEA.

(4) User Interfaces and Tools

EPA currently maintains a Web interface that allows access to IDEA for users who do not know the IDEA retrieval (keyword) language or the structure of the data in the source databases. This interface allows users to construct a query using a set of menus and pick lists, submit the query to the IDEA on the mainframe for retrieval and report generation, and view reports. The software requires ongoing maintenance and fixes are occasionally required as bugs are identified. The contractor shall continue to provide maintenance of this interface and make any enhancements or changes determined necessary by EPA. Communication between EPA and the contractor is required on a regular basis to assure that requirements are interpreted and understood. The contractor shall fulfill tasks in this area through a comprehensive understanding of the data and the "keyword" language for constructing queries in the mainframe environment.

As the Agency moves further to the use of the Internet to support access to data, IDEA will continue development of Web interfaces for use by EPA staff, the states, and the public. These interfaces will connect and communicate with the IDEA server on the mainframe to formulate and send user queries and receive and display data. The contractor shall be involved in all phases of Web interface design, protocols, communication interfaces, report output and display which shall require knowledge and experience in the Internet/Web environment, including XML. The contractor shall become familiar with EPA fire walls and security protocols imposed in the communication between the user's terminal and EPA servers, and incorporate the required security in the design and implementation of Intranet and Internet interfaces and communication.

(5) Identification of Data and Methods for Analysis, including Risk Assessment and Modeling

IDEA's strength is in the ability to perform analyses, incorporating varying sets of data and employing a variety of programs and software tools, including data normalization, graphics, mapping, and risk analyses to enhance the users ability to build queries and obtain reports. Tasks will be issued that shall require the contractor to further enhance existing tools and to develop new ones, such as statistics for program review by EPA management, measures reporting, contextual data that portraits the significance of data for different groupings or categories, and trend analysis.

Monthly Reporting Requirement

Monthly Progress Report - in Final form due by 15th of the following month to the EPA Project Officer and the EPA Contracting Officer in electronic format (Wordperfect, MS Word, PDF, or Excell spreadsheet.